09/980810
WO 00/77249
PCT/FR00/01617
JC13 Rec'd PCT/PTO 06 DEC 2001

2|prts

METHOD OF PRODUCTION OF COMPLEX MIXTURES OF cDNA AND APPLICATIONS OF THESE MIXTURES FOR THE ANALYSIS OF GENE EXPRESSION

ins A1

The invention relates to a method for the production of complex mixtures of cDNA and the applications of the said mixtures, in particular as probes for studying profiles of gene expression in a tissue or cells of animal, plant or microbial origin.

The methods that have been developed for collecting the expression profiles of a large number of genes are based on the hybridization of complex cDNA probes derived from messenger RNAs (mRNA's) on various substrates carrying either cDNA clones, or specific oligonucleotides of thousands of genes.

Obtaining cDNA's by reverse transcription (RT) of mRNA's comprises, conventionally, the use of, as primers, either oligo (dT), or of randomly synthesized oligonucleotides, or of specific oligonucleotides of the genes studied. In the first case, the oligo(dT) is attached to the poly(A) tail of the mRNA's (which are characteristic of the 3'OH end of eukaryotic messenger RNAs) and reverse transcriptase extends the oligo(dT) primer in the direction of the 5'P end of the mRNA's. In the second case, the primer, consisting of oligonucleotides synthesized at random, is attached randomly on the whole length of the mRNA and elongation of the cDNA is effected in the same conditions as those described above. In the 3rd case,

attachment is effected specifically on the genes being studied, and elongation is effected as described previously.

Using these cDNA's as probes, it is a question of simultaneously obtaining several thousands of hybridization signals which reflect biological situations and their dynamic variations for studies of physiology, pathology, or pharmacology, as well as for comparative studies of various model organisms.

It will be appreciated that the reliability of these studies depends on the nature of the cDNA mixture produced, which must reflect the complexity of the mRNA population regardless of the level of abundance of the molecular species of which it is constituted.

However, the technique cited above, which is the one most commonly used in laboratories, and the kits that are being marketed for cDNA synthesis by the two types of reaction indicated above, do not permit satisfactory transcription of the whole of an mRNA population, and especially of the mRNA's that are poorly represented.

In the standard technique of reverse transcription, each molecule of reverse transcriptase initiates the synthesis of a complementary chain of an mRNA, then stops spontaneously after elongation of several hundred nucleotides, and detaches itself from the chain in the course of synthesis. It then attaches itself to another chain in the course of synthesis and continues its elongation. This has the effect of favouring the most abundant mRNA's in the starting sample.

A method of gene cloning in which the mRNA corresponding to the gene is present in very small proportions in a mixture of mRNA is described in patent US 4 738928.

A short amino acid sequence of the required peptide (5 to 25 amino acids) must be known. This method then envisages preparing a complementary primer which is used in a reverse transcription process in an mRNA mixture containing the target. Chain terminators are used for stopping elongation.

In the article of Koch G. and Kant A. in Nucleic Acids Research, Vol. 18, No. 10, 1990, p. 3063-3065, the authors undertake the synthesis, in the conventional manner, of cDNA starting from genomic RNA, employing RT in a method of chain termination by dideoxy.

The inventors found that this technique could be improved by controlling transcription.

The aim of the invention is therefore to provide a method of producing very reliable and reproducible cDNA mixtures.

It also relates to these mixtures as such and their uses especially as hybridization probes.

The method according to the invention, for the production of a complex mixture of cDNA by reverse transcription of mRNA from tissues or cells, is characterized by the addition of elongation terminators in the reaction mixture, recovery of the cDNA mixture formed, preferably followed by its purification.

Surprisingly, the addition of elongation terminators has the effect of preventing the re-initiation mentioned above, and the molecules of reverse transcriptase then initiate the synthesis of complementary chains of the mRNA molecules that are poorly represented (the least abundant). As a result, the complex mixtures of cDNA produced in the presence of elongation terminators represent all of the starting mRNA, including the mRNA's present in small amounts.

The mRNA employed in the method of the invention is obtained from cells in culture or from samples or from tissue, and can be of any origin: animal, plant or microbial.

An elongation terminator that is widely used consists of dideoxynucleotides.

The reaction of reverse transcription is carried out in particular in accordance with the usual techniques.

Synthetic oligonucleotides, such as are employed conventionally in RT techniques, will be used advantageously as elongation primers. Oligonucleotides that are suitable for implementation of the method of the invention include hexamers or oligonucleotides synthesized at random. Labelling means are preferably added to the reaction medium, for example radioactive elements, fluorescent, luminescent or colorimetric agents, to provide labelled cDNA's for subsequent applications.

In general, the method according to the invention is applicable for all operations of production of cDNA by RT starting from the mRNA from a tissue or from a cell, of whatever origin.

It has the advantage of being highly reproducible, reliable, efficient, and does not give rise to a notable extra cost. The yield of the transcription reaction can reach 90% or even exceed this value, whereas it generally only reaches about 30% in procedures based on the usual conditions.

As shown in the examples, this method has the advantage of making it possible to study a very large number of genes and their levels of expression, whatever the species or tissue studied, as it is able to reveal genes that are poorly expressed.

The levels of gene expression can then be determined very reliably by counting cDNA's that have been cloned and sequenced, for example by the SAGE method or partial sequencing of cDNA libraries.

The invention also relates to kits for the application of the aforementioned method for purposes of synthesis of the said cDNA mixtures. These kits are characterized in that they include, in addition, reagents for carrying out reverse transcription, elongation terminators, in particular dideoxynucleotides, and instructions for use.

The invention relates to complex cDNA mixtures, as novel products, such as are obtained by the method defined above and using the said kits if necessary.

These mixtures are characterized in that they reliably reflect the state of transcription of a tissue or of cells, i.e. the number and the level of gene expression.

The said mixtures therefore constitute high-quality copies of transcriptons and thus make it possible to improve, in hybridization experiments, the performance of complex cDNA probes produced starting from these mixtures.

The invention therefore also relates to the use of complex cDNA mixtures as hybridization probes.

Advantageously, these probes make it possible to detect the expression of a large number of genes by greatly improving the capacity for detecting the activity of poorly expressed genes.

The quality of the cDNA's of the complex mixture, as probes, was tested by synthesizing complex probes starting from messenger RNAs from various tissues.

By using high-density filters, which make it possible to study the hybridization, simultaneously, of a large number of clones, the inventors demonstrated that the complex mixtures according to the invention made it possible to detect, on the filters, a number of clones 50 to 150% greater than is observed with the cDNA mixtures employed until now, the majority of the clones thus identified being clones corresponding to poorly expressed genes.

The invention therefore also relates to a method of studying the profile of gene expression in a tissue or in cells.

The said method comprises bringing the labelled cDNA mixtures defined above into contact with the DNA to be investigated (cDNA, or specific oligonucleotides of DNAs), in conditions permitting the hybridization of the complementary sequences when they are present.

The procedure is carried out in stringent or in non-stringent conditions, with substrates on which the DNAs to be studied are deposited.

These supports can be Nylon® filters, or other supports as well, including sheets of glass.

Suitable conditions correspond to the use of a temperature of 68°C for 2 hours with the same solution with 20×10^6 cpm of probe.

The filters are washed, and then the hybrids that have formed are revealed.

The invention thus provides means for studying gene expression and the variations of this expression as a function of certain factors.

Accordingly, it makes it possible to identify targets at the therapeutic level for drug development.

Examples that may be given are applications in oncology, chiefly in cancers of the colon, for identifying the genes that are altered in the cancerous cells relative to the healthy cells and for determining potential targets for development of therapies; in neuromuscular diseases for identifications of the aforementioned type in muscle, for example in Duchenne's muscular dystrophy (DMD); for carrying out studies on muscles in conditions of weightlessness, or in neurodegenerative diseases (Parkinson's disease or amyotrophic lateral sclerosis (ALS)).

Other applications in cattle and goats have the purpose of studying the modified genes in the mammary gland during gestation and lactation, in order to find targets for improving milk quality and quantity.

Other characteristics and advantages of the invention will be apparent from the examples given below, in which reference is made to Figs. 1 and 2, showing respectively, according to the four conditions of initiation of RT,

- Figure 1, the ratio between the size of the inserts in kb and the rates of hybridization,

- Figure 2A, the hybridization of high-density filters with complex cDNA probes derived from poly(A)$^-$ mRNA from murine skeletal muscle, and

- Figure 2B, the distribution of the clones according to type of hybridization.

Example 1: Production of a complex mixture of cDNA from mRNA of murine skeletal muscle.

The results of 5 series of experiments are given.

Each time, a complex mixture of cDNA is synthesized by reverse transcription of 500 ng of poly(A)$^-$ mRNA from murine skeletal muscle. The reaction is carried out using the SuperScript® pre-amplification system for synthesizing the first strand of cDNA (Life Technologies SARL, Cergy Pontoise, France), following the manufacturer's recommendations. The following are used: 500 ng of hexamers or 500 ng of oligo(dT) primers, 50 µCi of [α-^{33}P]dATP, 3000 Ci/mmol (Amersham France, S.A., Les Ulis, France) and 500 µM of d(T, C, G)TP (Pharmacia Biotech, Orsay, France), in a final volume of 50 µl.

The complex mixture obtained is purified on a column of Sephadex G-50® (Quick Spin®, Boehringer Mannheim, France S.A., Meylan, France).

The amount of radioactivity before and after purification is determined using a scintillation counter (Beckman Instruments France S.A., Gagny, France) for calculating the specific activity of the complex cDNA mixture (cpm/μg of cDNA) and the cDNA yield in the synthesis (%). The cDNA yield in the synthesis is calculated using a theoretical maximum of cDNA synthesis equal to 20 ng, a value that arises from the limiting concentration of dATP in the reaction (0.3 μM).

The following table gives the ratio between the conditions of initiation of the RT reaction, the specific activity of the complex cDNA mixtures (SA, cpm/μg of cDNA) and the cDNA yield in synthesis (% of the theoretical maximum).

Oligo(dT)		Oligo (dT) + ddTTP		Hexamers		Hexamers + ddTTP	
SA	Yield	SA	Yield	SA	Yield	SA	Yield
*3.10^9	44	6.10^9	28	5.10^9	1	5.10^9	100
3.10^9	15	6.10^9	31	5.10^9	21	3.10^9	100
2.10^9	17	↓5.10^9	38	5.10^9	19	5.10^9	95
4.10^9	42	5.10^9	39	*3.10^9	24	*5.10^9	89
4.10^9	89	5.10^9	39	3.10^9	9	5.10^9	79

"*" corresponds to the probes used for the hybridizations of filters shown in Figure 2.

Examination of this table shows that the specific activity of the complex cDNA mixtures is in the range from 2 to 6 x 10^9 cpm/μg regardless of the conditions of RT initiation. Using oligo(dT) primers or hexamers, the yield of the RT reaction, which ranges from 15 to 89% and from 1 to 24% respectively, seems very variable.

On the other hand, in the presence of dideoxynucleotides, the average yield of the reaction, equal to about 35% using oligo(dT) and to about 95% using hexamers, seems much more reproducible.

The data in this table also shows that the incorporation of dideoxynucleotides in the reaction mixture improves the efficiency of cDNA synthesis initiated with hexamers by a factor of 5 on average with all the values above 79%. Similar results are obtained working with 50 and 5 μM of ddTTP (i.e. 1/10 or 1/100 of dNTP concentration) and with ddCTP instead of ddTTP.

<u>Example 2</u>: Use of the complex cDNA mixture from Example 1 as a probe.

The purified complex cDNA mixture is used for purposes of hybridization on high-density Nylon® filters (Hybond-N+, Amersham France S.A., France), comprising the products of PCR amplification of the inserts of 1339 DNA clones from a bank of human skeletal muscle.

The filters are pre-hybridized at 68°C for 30 minutes in an ExpressHyb® hybridization solution (Clontech Inc., Palo Alto, CA, USA). Hybridization is carried out at 68°C for 2 hours in the same solution with 20 x 10^6 cpm of probe. The filters are washed twice at room temperature for 30 minutes in SSC 1X/0.1% of SDS, and twice for 30 minutes in SSC 0.1X/0.1% of SDS.

These non-stringent conditions permit optimum hybridization between the murine cDNA probes and the human cDNA targets. The filters are then exposed to phosphorus screens for 16 hours (Molecular Dynamics S.A., Paris, France).

For each filter, the 1339 hybridization signals are identified and quantified using software specially designed for this application (XdotReader, Cose, France). For each clone, the value of the rate of hybridization is calculated as described in Piétu et al., Genome Research, 1996, 6:492-503, and standardized by dividing by the mean of all the intensity values on each filter.

The use of hexamers and/or dideoxynucleotides in the RT reaction does not introduce any deviation in the hybridization of cDNA inserts of different lengths, as is shown by the absence of overall correlation between signal intensity and the length of the cDNA inserts ($r \leq 0.05$ with the four conditions of initiation tested, presented in Figure 1).

Figure 2A illustrates the qualitative aspects of the hybridization of high-density filters with the complex cDNA probes obtained in each condition tested.

Analysis of the hybridization signals enables us to assign intensity values to the 539 (40%), 451 (33%), 797 (60%) and 1122 (83%) clones, which differ from the background noise when the probe is primed with, respectively, oligo(dT), hexamers, oligo(dT) + ddTTP and hexamers + ddTTP.

These results show an overall increase of the clones detected in the presence of dideoxynucleotides: +47% and +148% using oligo(dT) primers or hexamers, respectively.

Figure 2B shows quantitative analysis of the values of hybridization rate distributed in four representative categories: background noise (B), weak (f), medium (M) : 2 times the weak value, and strong (F): 6 times the weak value. The symbols used have the following meanings: oligo(dT) primers: ■ , hexamers: ⊠ , oligo(dT) + ddTTP primers: □ , and hexamers + ddTTP: □ . The number of clones in each category of hybridization intensity is shown at the top of each histogram.

The number of clones with strong or medium values of hybridization rates shows limited variations when the four RT conditions are used.

On the contrary, the number of clones with weak intensity values is more than doubled when using the dideoxynucleotides with the oligo(dT) primers, and

multiplied nearly 18 times when they are used with hexamers. In these conditions the fraction of clones associated with the hybridization signals that cannot be distinguished from the background noise with a high reliability is limited to 16%, which illustrates the improved ability of the corresponding complex cDNA probe to reflect the least abundant transcripts.

Furthermore, the values of hybridization rates obtained with a complex cDNA probe produced with the hexamers and the dideoxynucleotides in 4 hybridizations of the same high-density filter differs by -10%.